JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 14, 153-160 (1966)

A Canonical Basis for the Matrix Transformation $X \to AXB$

ANTHONY TRAMPUS

Oceanics Division, Interstate Electronics Corporation,
Anaheim, California

Submitted by Ky Fan

1. INTRODUCTION

Let $\mathfrak{F}_{m,n}$ be the vector space of all m by n matrices over an algebraically closed field $\mathfrak{F}$. For any m by m matrix A and n by n matrix B with elements in $\mathfrak{F}$ we define linear transformations A^+ and B^- of $\mathfrak{F}_{m,n}$ by $A^+X = AX$ and $B^-X = XB$, respectively. The linear transformation A^+B^- may be represented by the mn by mn matrix $A \otimes B^T$ acting on the mn-dimensional vector space $\mathfrak{F}_{mn}$ (see Mac Duffee [1]) where $\otimes$ denotes tensor product (also known as Kronecker product or direct product) and B^T is the transpose of B. The present note establishes a basis of $\mathfrak{F}_{m,n}$ with respect to which the matrix representing A^+B^- is in Jordan canonical form. This basis is made up of matrices which are certain linear combinations of tensor products of vectors from the canonical bases belonging to A and B^T, in that order; it is much more complicated than the previously established canonical basis belonging to $A^+ + B^-$ in [2]. An immediate consequence of our result is the determination of the elementary divisors of A^+B^-, a result originally obtained by Roth [3] employing special matrix methods which are not appropriate for obtaining a canonical basis.

This note also covers the matrix $A \otimes B^T + \rho A \otimes I_n + \sigma I_m \otimes B^T$, where ρ and σ are arbitrary elements of $\mathfrak{F}$ and I_p is the identity p by p matrix, $p = m$ and n, since this matrix has the same canonical basis as $(A + \sigma I_m) \otimes (B + \rho I_n)^T$. A special case of the matrix

$$A \otimes B^T + \rho A \otimes I_n + \sigma I_m \otimes B^T$$

arises in connection with a finite difference approximation to the Poisson equation (see Lynch, Rice, and Thomas [4]).

The explicit determination of a canonical basis appears to be a new result. It should be noted that eigenvectors and generalized eigenvectors of a matrix of order mn are obtained from eigenvectors and generalized eigenvectors of

409/14/1-11

matrices of lower order, m and n. The previously determined elementary divisors merely indicate "how many" eigenvectors of various grades belong to the transformation. Our study indicates that if one wants to treat matrices of the form $\sum \gamma_{ij} A^i \otimes B^j$ it may be advantageous to deal instead with the transformation $\sum \gamma_{ij} (A^+)^i (B^-)^j$.

2. A Canonical Basis of $\mathfrak{M}$ Belonging to A^+B^-

We use the notation and terminology introduced in [2]. Let α and β be eigenvalues and let $(\lambda - \alpha)^e$ and $(\lambda - \beta)^f$ be elementary divisors of A and B, respectively. Let $x_e \in \mathfrak{F}_{m,1}$ and $y_f \in \mathfrak{F}_{n,1}$ be eigenvectors of A and B^T of grade e and f corresponding to eigenvalues α and β, respectively. We define

$$x_i = (A - \alpha I_m)^{e-i} x_e , \qquad i = 1, 2, \cdots, e \qquad (x_0 = 0)$$

$$y_j = (B^T - \beta I_n)^{f-j} y_f , \qquad j = 1, 2, \cdots, f \qquad (y_0 = 0).$$

We denote by $\mathfrak{M}$ the linear manifold spanned by the ef linearly independent matrices $\{x_i y_j^T\}$,

$$\mathfrak{M} = [x_1 y_1^T, x_1 y_2^T, \cdots, x_1 y_f^T, x_2 y_1^T, x_2 y_2^T, \cdots, x_2 y_f^T, \cdots, x_e y_1^T, x_e y_2^T, \cdots, x_e y_f^T] \subseteq \mathfrak{F}_{m,n} .$$

The subspace $\mathfrak{M}$ is invariant under A^+B^- since

$$\begin{aligned} A^+B^-(x_i y_j^T) &= (Ax_i)(B^T y_j)^T = (\alpha x_i + x_{i-1})(\beta y_j + y_{j-1})^T \\ &= \alpha\beta(x_i y_j^T) + \alpha x_i y_{j-1}^T + \beta x_{i-1} y_j^T, \quad \text{with} \quad x_0 = 0, \quad y_0 = 0, \\ &\qquad i = 1, 2, \cdots, e, \qquad j = 1, 2, \cdots, f. \end{aligned}$$

Clearly, $\mathfrak{F}_{m,n}$ is a direct sum of subspaces $\mathfrak{M}, \mathfrak{N}, \cdots$, invariant under A^+B^- obtained in the above manner by considering all elementary divisors of A and B and suitably associated eigenvectors. To obtain a canonical basis of $\mathfrak{F}_{m,n}$ it therefore suffices to consider only the case in which A and B each have a single elementary divisor.

Lemma. *The matrices $x_i y_j^T$ are eigenvectors of A^+B^- of grade k corresponding to the eigenvalue $\alpha\beta$, where (a) if $\alpha\beta \neq 0$, $k = i + j - 1$, (b) if $\alpha \neq 0$ and $\beta = 0$, $k = j$, (c) if $\alpha = 0$ and $\beta \neq 0$, $k = i$, and (d) if $\alpha = 0$ and $\beta = 0$, $k = \min(i, j)$, $i = 1, 2, \cdots, e$, $j = 1, 2, \cdots, f$.*

PROOF. Let

$$\begin{aligned}(x_iy_j^T)_{(t)} &= (A^+B^- - \alpha\beta I_m{}^+)^t (x_iy_j^T) \\ &= [(A - \alpha I_m)^+ (B - \beta I_n)^- + \alpha(B - \beta I_n)^- + \beta(A - \alpha I_m)^+]^t (x_iy_j^T) \\ &= \sum_{s=0}^{t} \sum_{\substack{r=0 \\ r+s \leqslant t}}^{t} \frac{t!}{r!s!(t-r-s)!} \quad \alpha^r\beta^s[(A - \alpha I_m)^{t-r} x_i] [(B^T - \beta I_n)^{t-s} y_j]^T\end{aligned}$$

If $\alpha\beta \neq 0$ and $t = i + j - 2$, then the r, s-term of the last sum is not zero if and only if $r \geqslant j - 1$ and $s \geqslant i - 1$. Since $r + s \leqslant i + j - 2$, we must have

$$(x_iy_j^T)_{(i+j-2)} = \frac{(i+j-2)!}{(i-1)!\,(j-1)!} \alpha^{j-1}\beta^{i-1}(x_1y_1^T) \neq 0$$

and $(x_iy_j^T)_{(k)} = 0$ for $k \geqslant i + j - 1$. This proves case (a). In case (b) we have

$$\begin{aligned}(x_iy_j^T)_{(k)} &= \sum_{r=0}^{k} \binom{k}{r} \alpha^r[(A - \alpha I_m)^{k-r} x_i] [(B^T - \beta I_n)^k y_j]^T \\ &= \sum_{r=0}^{j-1} \binom{j-1}{r} \alpha^r[(A - \alpha I_m)^{j-1-r} x_i]\, y_1^T \neq 0, \qquad \text{if} \qquad k = j - 1 \\ &= 0, \qquad \text{if} \qquad k = j.\end{aligned}$$

The proof of case (c) is similar to that of (b). In case (d) we have

$$\begin{aligned}(x_iy_j^T)_{(k)} &= [(A - \alpha I_m)^k x_i] [(B^T - \beta I_n)^k y_j]^T \\ &= x_{i-v+1}y_{j-v+1}^T \neq 0, \qquad \text{if} \qquad k = v - 1 \\ &= 0, \qquad \text{if} \qquad k \geqslant v = \min(i, j).\end{aligned}$$

This completes the proof of the lemma.

In what follows, we shall make use of the notation and elementary properties of factorial polynomials, binomial coefficients and finite differences. We proceed to list all the facts we shall need.

$$s^{(t)} = s(s-1)(s-2) \cdots (s-t+1), \qquad s^{(0)} = 1$$

$$s^{(t+u)} = s^{(t)}(s-t)^{(u)} = s^{(u)}(s-u)^{(t)}$$

$$\binom{s}{t} = \frac{s^{(t)}}{t!}$$

$$\Delta^t s_j = \sum_{i=0}^{t} (-1)^i \binom{t}{i} s_{t-i+j} = (-1)^t \Delta^t r_j \,, \qquad r_{t-i+j} = s_{i+j}$$

$$\Delta^t(r_j s_j) = \sum_{i=0}^{t} \binom{t}{i} (\Delta^{t-i} r_{j+i}) (\Delta^i s_j) = \sum_{i=0}^{t} \binom{t}{i} (\Delta^{t-i} r_j) (\Delta^i s_{j+t-i})$$

$$\begin{aligned} \Delta^k (s+j)^{(u)} (t-j)^{(v)} &= 0, && \text{if} \quad k > u+v \\ &= (-1)^v (u+v)!, && \text{if} \quad k = u+v \end{aligned}$$

$$\binom{s-t}{k} = \sum_{i=0}^{k} (-1)^i \binom{s}{k-i} \binom{t-1+i}{t-1} \qquad \text{(see [5])}$$

$$\frac{(s+j)^{(t+v-i)}}{(v-i)!} = (s+j)^{(t)} \Delta^i \binom{s-t+j}{v}.$$

Here as well as in the following, the difference Δ is with respect to the integer variable j.

THEOREM. *We define a set $\{X_{ku}\}$ consisting of ef matrices according to the following prescription.*

(a) *If $\alpha\beta \neq 0$, let $k_u = e + f - 2u + 1$, $\mu = \min(e, f)$, and*

$$X_{k_1 1} = x_e y_f^T$$

$$X_{k_u u} = \sum_{i=0}^{u-2} \binom{u-1}{i} X_{k_u u}^{(i)} \,, \; u = 2, 3, \cdots, \mu$$

with

$$X_{k_u u}^{(i)} = \sum_{j=0}^{u-1-i} (-1)^j \binom{u-1-i}{j} (e-1-j)^{(u-1-j)} (f-u+j)^{(j)} \times \alpha^{u-1-i-j} \beta^j x_{e-i-j} y_{f-u+1+j}^T$$

(b) *If $\alpha \neq 0$ and $\beta = 0$, let $k_u = f$, $\mu = e$, and*

$$X_{k_u u} = x_u y_f^T, \qquad u = 1, 2, \cdots, \mu.$$

(c) *If $\alpha = 0$ and $\beta \neq 0$, let $k_u = e$, $\mu = f$, and*

$$X_{k_u u} = x_e y_u^T, \qquad u = 1, 2, \cdots, \mu.$$

(d) *If* $\alpha = 0$ *and* $\beta = 0$, *let* $\mu = e + f - 1$, *and*

$$\begin{aligned}
X_{k_u u} &= x_e y^T_{e+u-1}, \quad \text{if} \quad e = \min(e,f) \Big\} \quad k_u = \min(e,f) \\
&= x_{f+u-1} y_f{}^T, \quad \text{if} \quad f = \min(e,f) \Big\} \quad u = 1, 2, \cdots, |e - f| + 1 \\
&= x_{k_u} y_f{}^T, \qquad k_u = \tfrac{1}{2}(e + f - u), \\
&\qquad\qquad u = |e - f| + 2, |e - f| + 4, \cdots, e + f - 2 \\
&= x_e y^T_{k_u}, \qquad k_u = \tfrac{1}{2}(e + f + 1 - u), \\
&\qquad\qquad u = |e - f| + 3, |e - f| + 5, \cdots, e + f - 1.
\end{aligned}$$

Then $\{X_{ku}\} = \{X_{11}, X_{21}, \cdots, X_{k_1 1}, X_{22}, \cdots, X_{k_2 2}, \cdots, X_{1\mu}, X_{2\mu}, \cdots, X_{k_\mu \mu}\}$ *is a canonical basis of* $\mathfrak{M}$ *with respect to which the matrix representing* A^+B^- *is in Jordan canonical form.*

PROOF. We first show that $X_{k_u u}$ is an eigenvector of A^+B^- of grade k_u corresponding to the eigenvalue $\alpha\beta$. The lemma already gives this result in cases (b), (c), (d), and for $u = 1$ in case (a). Suppose $\alpha\beta \neq 0$ and $u > 1$. Then

$$(A^+B^- - \alpha\beta I_m{}^+)^{k_u} X_{k_u u}$$

$$= \sum_{\substack{s=0 \\ r+s}}^{k_u} \sum_{\substack{r=0 \\ \leqslant k_u}}^{k_u} \frac{(k_u)!}{r!s!(k_u - r - s)!} \alpha^r \beta^s (A - \alpha I_m)^{k_u - r} X_{k_u u}(B - \beta I_n)^{k_u - s}$$

with

$$(A - \alpha I_m)^{k_u - r} X_{k_u u}(B - \beta I_n)^{k_u - s}$$

$$= \sum_{i=0}^{u-2} \sum_{j=0}^{u-1-i} (-1)^j \binom{u-1}{i} \binom{u-1-i}{j} (e - 1 - j)^{(u-1-j)} (f - u + j)^{(j)}$$

$$\times \alpha^{u-1-i-j} \beta^j \, [(A - \alpha I_m)^{k_u - r} x_{e-i-j}] \, [(B^T - \beta I_n)^{k_u - s} y_{f-u+1+j}]^T.$$

Since $X_{k_u u}$ is a linear combination of only those $x_i y_j{}^T$ which are of grade $i + j - 1 = e + f - u, e + f - u - 1, \cdots, e + f - 2u + 2$, we see that the result of the above expansion will be a linear combination of only those $x_i y_j{}^T$

which are of grade $i + j - 1 = u - 1, u - 2, \cdots, 1$. After some simplifications we find that

$$(A^+B^- - \alpha\beta I_m{}^+)^{k_u} X_{k_u u}$$

$$= \frac{(k_u)!}{(e-u)!\,(f-u)!} \sum_{v=0}^{u-2} \sum_{t=0}^{u-2-v} \sum_{i=0}^{v} \sum_{j=0}^{u-1-i} \frac{(-1)^j}{(v-i)!} \binom{u-1}{i} \binom{u-1-i}{j}$$

$$\times (e-1-j)^{(u-t-2)} (f-u+j)^{(t+v-i)} \alpha^{f-1-v-t} \beta^{e-u+1+t} x_{u-1-v-t} y_{1+t}^T$$

$$= \frac{(-1)^{u-1}(k_u)!}{(e-u)!\,(f-u)!} \sum_{v=0}^{u-2} \sum_{t=0}^{u-2-v} \sum_{i=0}^{v} (-1)^i \binom{u-1}{i} \Big\{ \Delta^{u-1-i} \Big[(e-1-j)^{(u-t-2)}$$

$$\times (f-u+j)^{(t)} \Delta^i \binom{f-u-t+j}{v} \Big] \Big\}_{j=0} \alpha^{f-1-v-t} \beta^{e-u+1+t} x_{u-v-1-t} y_{1+t}^T.$$

We expand the coefficients employing the rule for differencing a product to get

$$\sum_{i=0}^{v} (-1)^i \binom{u-1}{i} \Big\{ \Delta^{u-1-i} \Big[(e-1-j)^{(u-t-2)}$$

$$\times (f-u+j)^{(t)} \Delta^i \binom{f-u-t+j}{v} \Big] \Big\}$$

$$= \sum_{k=0}^{u-1} \Big[\Delta^{u-1-k} \binom{f-u-t+k+j}{v} \Big] \big[\Delta^k (e-1-j)^{(u-t-2)} (f-u+j)^{(t)} \big]$$

$$\times \Big[(-1)^v \sum_{i=0}^{v} (-1)^i \binom{u-1}{v-i} \binom{u-v-1+i}{k} \Big]$$

$$= 0$$

since

$$\Delta^{u-1-k} \binom{f-u-t+k+j}{v} = 0, \qquad k = 0, 1, \cdots, u-v-2,$$

$$\sum_{i=0}^{v} (-1)^i \binom{u-1}{v-i} \binom{u-v-1+i}{k} = (-1)^s \binom{v-1-s}{v-s} = 0,$$

$$k = u - v - 1 + s, \qquad s = 0, 1, 2, \cdots, v-1,$$

and

$$\Delta^k (e-1-j)^{(u-t-2)} (f-u+j)^{(t)} = 0, \qquad \text{if} \qquad k = u-1.$$

Since this is true for $0 \leqslant t \leqslant u - 2 - v$ and $0 \leqslant v \leqslant u - 2$, we have

$$(A^+B^- - \alpha\beta I_m{}^+)^{k_u} X_{k_u u} = 0.$$

Next we note that

$$X_{ku} = (A^+B^- - \alpha\beta I_m{}^+)^{k_u - k} X_{k_u u} = X_{ku}^{(0)} + Y_{ku}, \qquad k = 1, 2, \cdots, k_u$$
$$u = 1, 2, \cdots, \mu,$$

where $\{X_{ku}^{(0)}\}$ is a canonical basis of $\mathfrak{M}$ belonging to $\beta A^+ + \alpha B^-$ ([2]); $X_{ku}^{(0)}$ is a linear combination of $x_i y_j{}^T$ with $i + j = u + k$ and Y_{ku} is a linear combination of $x_i y_j{}^T$ with $i + j < u + k$. Hence $X_{1u} \neq 0$, so that $X_{k_u u}$ is an eigenvector of A^+B^- of grade k_u corresponding to the eigenvalue $\alpha\beta$, $u = 1, 2, \cdots, \mu$. The linear independence of the X_{ku} follows on observing that no matrix X_{ku} is a linear combination of the preceding ones in the sequence

$$\begin{array}{l}
X_{e+f-1,1} \\
X_{e+f-2,1} \\
X_{e+f-3,1}, X_{e+f-3,2} \\
X_{e+f-4,1}, X_{e+f-4,2} \\
\vdots \\
X_{e+f-2\mu+1,1}, X_{e+f-2\mu+1,2}, \cdots, X_{e+f-2\mu+1,} \\
X_{e+f-2\mu,1}, X_{e+f-2\mu,2}, \cdots, X_{e+f-2\mu,\mu} \\
\vdots \\
X_{1,1}, X_{1,2}, \cdots, X_{1,\mu}.
\end{array}$$

This completes the proof for case (a).

In case (b)

$$X_{ku} = [(A - \alpha I_m)^+ (B - \beta I_n)^- + \alpha(B - \beta I_n)^-]^{f-k} x_u y_f{}^T$$

$$= \sum_{i=0}^{f-k} \binom{f-k}{i} \alpha^i [(A - \alpha I_m)^{f-k-i} x_u] [(B^T - \beta I_n)^{f-k} x_f]^T$$

$$= \sum_{\substack{i \geqslant f-k-u+1 \\ i \geqslant 0}}^{f-k} \binom{f-k}{i} \alpha^i x_{u-f+k+i} y_k{}^T, \qquad \begin{array}{ll} k = 1, 2, \cdots, & k_u = f; \\ u = 1, 2, \cdots, & \mu = e. \end{array}$$

The linear independence of the X_{ku} follows on observing that X_{ku} has nonzero components only in the directions $x_1 y_k{}^T$, $x_2 y_k{}^T$, $\cdots$, $x_u y_k{}^T$ and the fact that $\{x_i y_j{}^T\}$ is a basis of $\mathfrak{M}$.

The proof of linear independence is case (c) is similar to that of case (b). In case (d) it can be shown that $\{X_{ku}\} = \{x_i y_j\}$ and the linear independence follows. This completes the proof of the theorem.

As an immediate consequence of the above theorem, we find that a maximal set of linearly independent solutions of the matrix equation $AXB = \lambda X$ in $\mathfrak{M}$ is $X_{11}, X_{12}, \cdots, X_{1\mu}$, if $\lambda = \alpha\beta$; for other values of λ no solution exists in $\mathfrak{M}$. The problem of finding all automorphic transformations of a given matrix is equivalent to solving a special case of the equation $AXB = \lambda X^T$, $\lambda \neq 0$ (see Wedderburn [6]). This in turn is equivalent to solving the equation $(AB^T)\, Y(A^TB) = \lambda^2 Y$ and setting $X = Y^T$.

3. Elementary Divisors of A^+B^-

The following result, originally obtained by Roth [3], is an immediate consequence of our theorem in view of the direct sum decomposition of $\mathfrak{F}_{m,n}$.

Corollary. *If A and B have elementary divisors $(\lambda - \alpha_i)^{e_i}$, $i = 1, 2, \cdots, p$, $\sum_{i=1}^p e_i = m$ and $(\lambda - \beta_j)^{f_j}$, $j = 1, 2, \cdots, q$, $\sum_{j=1}^q f_j = n$, respectively, then A^+B^- has elementary divisors*

(a) $(\lambda - \alpha_i\beta_j)^{e_i+f_i-2u+1}$, $u = 1, 2, \cdots, \mu_{ij} = \min(e_i, f_j)$, *if* $\alpha_i\beta_j \neq 0$,

(b) λ^{f_j} *repeated* e_i *times, if* $\alpha_i \neq 0$, $\beta_j = 0$

(c) λ^{e_i} *repeated* f_j *times, if* $\alpha_i = 0$, $\beta_j \neq 0$

(d) $\lambda^{\mu_{ij}}$ *repeated* $|e_i - f_j| + 1$ *times*; $\lambda^{\mu_{ij}-1}, \lambda^{\mu_{ij}-2}, \cdots, \lambda$ *each repeated twice*, $\mu_{ij} = \min(e_i, f_j)$, *if* $\alpha_i = 0$, $\beta_j = 0$.

Proof. The exponents are merely the values of k_μ associated with α, β, e, and f in the theorem, since k_u is the order of a Jordan block matrix in the Jordan canonical form for A^+B^-.

References

1. C. C. MacDuffee. "The Theory of Matrices," Chaps. VII and VIII. Chelsea, New York, 1956.
2. A. Trampus. A canonical basis for the matrix transformation $X \to AX + XB$. *J. Math. Anal. Appl.* **14**, (1966).
3. W. E. Roth. On direct product matrices. *Bull. Am. Math. Soc.* **40** (1934), 461-468.
4. R. E. Lynch, J. R. Rice, and D. H. Thomas. Tensor product analysis of partial difference equations. *Bull. Am. Math. Soc.* **70** (1964), 378-384.
5. W. E. Milne. "Numerical Calculus," p. 154. Princeton Univ. Press, Princeton, New Jersey, 1949.
6. J. H. M. Wedderburn. "Lectures on Matrices," Chap. IX. Math. Soc. Colloq. Publ. 1934.